

```

1 2020-08-04
2 -----
3 Minu
4 ----
5     1. loop, if-else removal, map/filter
6     2. sampled using suffle.
7     Questions:
8     -----
9     3.
10
11 Astha
12 -----
13     1. Vectorizer and angle calculation done.
14     Questions:
15     2.
16
17 Shankar
18 -----
19     1. Indexed vectorizer done.
20     2. Environment conflicts.
21     3. Performance evaluation.
22 -----
23
24 Labin
25 -----
26 palindrome identification
27 -----
28 s1 = maaaam
29 s2 = madam
30
31     >> def [x1, (reverse s1)]
32     >> (= x1 s1)
33
34 [m, a, d, a, m ]
35
36 partition = [m, a]
37           = [a, m]
38
39           = [m, a, a]
40           = [a, a, m]
41
42 push => m, a, pop a, m
43
44     (a, pop, a)
45     (m, pop, m)
46     not-equal => false.
47
48     (= a a) each pop
49     (= a a)
50     (= m m)
51
52 -----
53 mid-position, start push, after, start pop.
54 -----
55 1M length.
56 -----
57 w1          w2
58 eat        vs  ate
59 feet       vs  foot
60
61 sim(w1, w2) = ???
62
63 edit distance
64 -----

```

```

65 insertion
66 deletion
67 replacement
68 transposition
69
70 eat -> ate => 4 character
71 eat => delete(e)
72 at => insert(e)
73
74 edit distance =>
75
76 lab => labin : i, n insert
77
78 1M words
79 -----
80 a-z 26 characters ,
81
82 lab => labing
83 3 len(incomplete word) => 6 dictionary
84 3 to 6 operations choose from 26 characters.
85 -----
86 based on character replacement, .....
87
88 -----
89 eat => ate
90
91 you(person) eat rice(food).
92 he(person) ate fruit(food).
93 .....
94
95 Suggestion problem
96 -----
97 Words are iid. (independent and identically distributed random variable)
98
99 And also,
100 the n-grams.
101 -----
102 Simplify the problem.
103
104
105 -----
106 n = 2, 3, 4, 5....
107 how do we choose it ?
108
109 shortest = 3
110 longest = 22
111
112 what happens
113 if n = 1, 2: over suggestion / over fitting
114
115 or n = 6, 7: under suggestion / under fitting
116 ???
117
118 -2d <=> +2d
119 -----
120 n 2, 3, 4, 5, 6
121 -----
122 We use multi-grams
123 n = [2, 3], [3, 4], [4, 5], [5, 6]
124 n = [2, 3, 4], [3, 4, 5], [4, 5, 6]
125 -----
126 Parameter estimation.
127 -----
128 Test data

```

```
129 -----
130 word1 => suggestion1{}
131 word2 => suggestion2{}
132 .
133 .
134 .
135 wordn => suggestionn{}
136 -----
137 accuracy??, precision?? recall ??
138 we chosse best result of n.
139 ???
140 -----
141 precision vs recall
142
143 if we increase precision, it will decrease recall and vice versa.
144 -----
145 Seen examples: Train
146 Unseen examples: Test
147
148
```